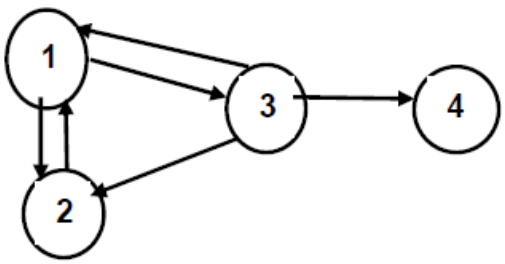
SVKM's
D. J. Sanghvi College of Engineering

Program: B.Tech in Electronics & Telecommunication Engg **Academic Year: 2022** **Duration: 3 hours**
Date: 06.01.2023
Time: 10:30 am to 01:30 pm
Subject: Big Data Analytics (Semester VII) **Marks: 75**

Question No.		Max. Marks
Q1 (a)	Describe any five characteristics of Big Data. Discuss a case study which applies these characteristics.	[07]
Q1 (b)	With suitable block diagram explain architecture of HDFS in Hadoop. Discuss role of Data node and Name node in HDFS. **OR** Explain Hadoop Ecosystem and briefly explain its components in four layers.	[08] [08]
Q2 (a)	Utilize CAP theorem in Amazon e-commerce application and show its usage. **OR** Apply CRUD operations using MongoDB with syntax and example.	[05] [05]
Q2 (b)	What is NoSQL? Differentiate NOSQL and SQL in detail.	[10]
Q3 (a)	What is node failure? Explain coping with node failure in MapReduce.	[05]
Q3 (b)	Explain various phases of Map-Reduce execution pipeline and role of combiner with suitable example. **OR** Write Map function, Reduce function and Pseudo code for "Selection", "Projection" .	[10] [10]
Q4 (a)	Illustrate with an example the application of the Apriori algorithm to a relatively simple case that generalizes to those used in practice. Show how to use the Apriori algorithm to generate frequent item sets and rules and to evaluate and visualize the rules. **OR** Compute the modified page rank of nodes with teleport factor 0.8. (Show two iterations). Does this network have dead ends and spider traps? 	[10] [10]

Q4 (b)	Explain any five applications of Nearest-neighbor search.	[05]
Q5 (a)	Describe components of apache spark in detail. Also list the features of Spark.	[08]
Q5 (b)	What are the problems related to Map Reduce data storage? How Apache Spark solves it using Resilient Distributed Dataset? Explain RDDs in Detail. **OR** What is transformation and actions in Apache Spark? Discuss various commands available for this activities in Apache Spark?	[07] [07]